

Lösningar till Tentamen 2003-04-15 Databasteknik - 1DL116

DatumTisdagen den 15 April, 2003
Tid09:00-14:00
Jourhavande lärare ...Kjell Orsborn, tel. 471 11 54 eller 070 425 06 91
HjälpmedelInga

Anvisningar:

- Läs igenom hela skrivningen och notera eventuella oklarheter innan du börjar lösa uppgifterna. Förutom anvisningarna på skrivningsomslaget så gäller följande:
 - Skriv tydligt och klart. Lösningar som inte går att läsa kan naturligtvis inte ge några poäng och oklara formuleringar kan dessutom misstolkas.
 - Antaganden utöver de som står i uppgiften måste anges. Gjorda antaganden får förstås inte förändra den givna uppgiften.
 - Skriv endast på en sida av papperet och använd ett nytt papper för varje uppgift för att underlätta rättning och minska risken för missförstånd.
- För godkänt krävs det cirka 50% av maxpoäng.

1. DB terminologi:

4p

Förklara kortfattade följande begrepp i databassammanhang (det går bra med definitioner där sådana finns).

- (a) metadata
- (b) kandidatnyckel
- (c) fullt funktionellt beroende
- (d) objektidentifierare (OID)

Svar:

- (a) metadata i databassammanhang är data om data, dvs metadata beskriver innehållet och strukturen hos databasen inklusive namn och struktur för filer, namn och datatyper för datafält och olika begränsningsvillkor för dessa.
- (b) en kandidatnyckel är varje minimal delmängd av en relations attribut som unikt kan identifiera alla tupler i relationen (notera att det kan finnas fler än en kandidatnyckel för samma relation).
- (c) fullt funktionellt beroende anger att för ett funktionellt beroende gäller att det inte finns någon delmängd attribut $A \subseteq X$ så att $(X - \{A\})Y$. Här gäller att R är ett relationsschema och $r(R)$ är en instans av schemat R med attributen A_1, \dots, A_n och $X, Y \subseteq \{A_1, \dots, A_n\}$. Ett funktionellt beroende så att X bestämmer Y , $X - Y$, existerar då om för varje par av tupler $t_1, t_2 \in r(R)$ och för alla $r(R)$ följande gäller: om $t_1[X] = t_2[X]$ så gäller att $t_1[Y] = t_2[Y]$. Alltså ett fullt funktionellt beroende är ett funktionellt beroende som inte innehåller något onödigt attribut i determinanten (vänsterledet i beroendet).
- (d) en objektidentifierare (OID) är en unik och ofta logisk och system-genererad identifierare som används för att unikt identifiera objekt under hela dess existens, samt för att hantera referenser mellan objekt.

2. Konceptuell modellering (Enhanced Entity-Relationship modellering): 4p

Vid specifikationen av en specialisering eller generalisering (eng. specialization/generalization) kan man definiera olika bivillkor (eng. constraints). Förklara i detta sammanhang kortfattat följande begrepp:

- (a) disjointness constraint
- (b) completeness constraint

Svar:

- (a) disjointness constraint innebär att subklasserna till en superklass kan specificeras till att vara *disjunkta*, dvs att en instans kan endast vara medlem av en av dessa subklasser, eller *överlappande* (eng. overlapping), dvs disjunktvillkoret krävs så att instanser tillåts vara medlemmar till fler än en av subklasserna.

- (b) completeness constraint specificerar klassificeringen till att vara *total* eller *partiell*. Om klassificeringen är total så måste instanserna till superklasserna vara medlem i någon subklass och för en partiell klassificering behöver ej instanserna till superklassen vara medlem av någon subklass.

Notera att villkoren a), b) är ortogonala, dvs de är oberoende av varann så att alla kombinationer av dessa kan existera.

3. Relationsalgebra: 4p

Skriv *relationsalgebrauttryck* för följande frågor:

- (a) **select** namn **from** person **where** person.avdnr = 5
(b) **select** namn, avdnamn **from** person p, avdelning a **where** p.avdnr=a.avdnr

4. Datamodeller: 4p

Förklara, och ge exempel på, vad som menas med begreppen *fysiskt* och *logiskt databeroende* som exempelvis hävdas att man kan uppnå med tre-nivå-arkitekturen.

Svar:

Fysiskt databeroende är förmågan att kunna förändra det interna schemat utan att behöva förändra det konceptuella (ej heller det externa) schemat. Ett exempel kan vara att man vill definiera ett index över ett attribut i en tabell för att snabba upp sökningen av vid frågor som inbegriper detta attribut. Dessa frågor behöver alltså inte definieras om för att utnyttja detta index.

Physical data independence The possibility to change the internal schema without influencing the conceptual schema. e.g. the effects of a physical reorganization of the database, such as adding an access path, is eliminated.

Logiskt databeroende The possibility to change the conceptual schema without influencing the external schemas (views). e.g. add another field to a conceptual schema.

5. Transaktionshantering: 4p

Beskriv de egenskaper som man ofta vill att transaktioner skall uppfylla i databassammanhang (ledning: ACID).

Svar:

To preserve the integrity of data, the DBMS must ensure ACID properties:

- Atomicity (atomic or indivisible): a logic processing unit (all operations of the transaction) is carried out in its whole or not at all.

- Consistency (preservation): a correct execution of a transaction in isolation should preserve the consistency of the database (from one consistent state to another).
- Isolation: Although multiple transactions may execute concurrently, each transaction must be unaware of other concurrently executing transactions. The updates of a transaction shall be isolated from other transactions until after the commit point.
- Durability (or permanency): If a transaction completes successfully, the changes it has made to the database must persist and should not be lost in a later system failure.

6. Återhämtning (eng. recovery): 4p

Beskriv kortfattat proceduren för återhämtning enligt modellen uppskjuten uppdatering (eng. deferred update) i en fleranvändarversion.

Svar: Återhämtning enligt modellen uppskjuten uppdatering (no-undo/redo):

1. Starta from the last record in the log file and traverse backwards. Create two lists:
 - C transactions that have reached their commit points
 - NC transactions that have not reached their commit points.
2. Start from the beginning of the log file and redo all (Write,T,...) for all transactions T in the list C.
3. Restart all transactions in the list NC. If the log file is long, step 2 will take long time. An improvement of this method is accomplished by introducing what's called check points.

7. Fysisk datalagring - filstrukturer: 4p

Beskriv kortfattat organisationen av och funktionen för hash-filer (svaret skall bland annat innefatta hur man återfinner (söker) en datapost för en specifik sökning där datafilen är en *hash-fil*)

Svar:

En hash-fil består av ett statiskt eller dynamiskt antal datablock som hanteras av olika typer av hashningstekniker. Hash-filer hanterar adressering av dataposter till datablock genom att applicera en hash-funktion till hash-fältet (dvs sökfältet) vilken returnerar adressen till ett datablock för insättning eller återsökning av dataposten. En vanlig form av hash-funktion har formen $h(f(p)) = f(p) \bmod M$, där hash-funktionen $h(f(p))$ tillhandahåller adressen för det datablock där dataposten p skall lagras genom att beräkna hashfältet $f(p)$ modulo (mod) antalet datablock M . Man hittar alltså var (i vilket block) en datapost finns för en specifik sökning genom att beräkna hash-funktionen för nyckeln som ger adressen till blocket.

8. Distribuerade databassystem: 4p

Förklara kortfattat följande begrepp inom distribuerade databaser:

- (a) fragmentering och fragmenteringstransparens (eng. fragmentation and fragmentation transparency).

- (b) replikering och replikeringstransparens (eng. replication and replication transparency).

Svar:

- (a) fragmentering : en uppdelning av relationerna i en databas i logiska enheter som kan lagras på olika noder i ett nätverk av databasservrar. De logiska enheterna kallas fragment. Tre olika sorter : horisontell (horizontal), vertikal (vertical) samt blandad (mixed/hybrid) fragmentering. Vid horisontell fragmentering skapas och distribueras delmängder av tuplerna i en relation. Vid vertikal fragmentering skapas och distribueras delmänder av attributen i en relation. Vertikal fragmentering måste genomföras så att varje fragment (mängd av ett antal attribut för relationen) innehåller primärnyckeln. Detta innebär att den ursprungliga relationen kan återskapas på ett korrekt sätt, dvs varje tupel i relationen kan identifieras och sättas ihop korrekt till ursprunglig form. Mixad fragmentering är en blandning av horisontell och vertikal fragmentering. Fragmenteringstransparens gör användaren omedveten om att en databas (dvs dess relationer) är fragmenterad.
- (b) replikering : Att lagra samma information (dvs en relation eller fragment) på flera noder i ett nätverk av databaseservrar. Extremfallen är ingen replikering alls där varje fragment lagras på endast en nod och fullständig replikering där hela databasen lagras på varje nod. Mellan extremfall erhålles partiell replikering där en del av fragmenten kan vara replikerade men andra inte. (Behövs ej: Mindre replikering minskar arbetsbördan vid uppdateringar men minskar också tillgängligheten. Mer replikering ökar arbetsbördan vid uppdateringar men ökar också tillgängligheten.) (Behövs ej: Ett replication schema beskriver var alla fragment är lagrade.) Replikeringstransparens gör användaren omedveten om att en databas (dvs dess fragment) är replikerad. (Ref kap 24.1.2, 24.2, 24.2.1-24.2.2 i vita Elmasri/Navathe - 3:e ed)

GLAD PÅSK!

/ Kjell